**CHAPITRE 3**

**STATISTIQUE DESCRIPTIVE A DEUX DIMENSIONS :**

**ETUDE D’UNE SERIE STATISTIQUE DOUBLE**

**Les objectifs de ce cours sont au nombre de Deux**.

**Premier Objectif**

Etudier l’existence de Corrélation entre deux caractères Quantitatifs.

**Second Objectif**

Etudier le modèle de Régression le plus Approprié pour expliquer la corrélation entre deux caractères Quantitatifs.

A.- **Pour répondre au Premier Objectif**,

On doit au préalable Ordonner la série double ; puis répondre à la question de corrélation pouvant exister entre les deux caractères quantitatifs.

**1.-Définitions de la série double**

**d1.**On appelle série statistique double une suite finie de couples (xi,yi)i=1,…n où :

xi est la mesure du caractère X sur l’individu n°i

et yi est la mesure du caractère Y sur l’individu n°i

**d2.** On appelle xi la modalité n°i du caractère X et

On appelle yj la modalité n°j du caractère Y.

**2.-Exemples**

**1**.- On donne pour chaque ville :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ville |  | Casa | Rabat | Marrakech | Tétouan | Agadir | Nador | Tiznit |
| Ensoleillement |  | 2072 | 1729 | 2763 | 1574 | 1685 | 1833 | 2790 |
| Température |  | 19.4 | 18.8 | 24.2 | 12.5 | 13.8 | 11.7 | 28.4 |

* Population ciblée : Les 7 villes étudiées
* Caractère X :

Le nombre moyen d’heures d’ensoleillement par année

* Caractère Y :

La température moyenne

**2**.- On donne l’évolution de l’espérance de vie au MAROC de 1990 à 1999 pour les femmes :

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Année | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| Espérance  de vie | 70.1 | 71.1 | 71.3 | 71.4 | 71.8 | 71.9 | 72.0 | 72.3 | 72.4 | 72.4 |

* Population ciblée : Les femmes du MAROC
* Caractère X : Année
* Caractère Y : L’espérance de vie : Le nombre d’années à vivre

**3.-** Chaquemois, une entreprise consacre :

* Caractère X : Une somme d’argent à des opérations publicitaires
* Caractère Y : Ventes réalisées.

Une étude est réalisée sur 8 mois a donnée les résultats suivants exprimées en millions de dirhams.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frais de Publicité | 0.24 | 0.30 | 0.25 | 0.32 | 0.35 | 0.20 | 0.18 | 0.30 |
| Ventes réalisées | 38 | 42 | 39 | 40 | 45 | 35 | 34 | 41 |

* Caractère X : somme d’argent à des opérations publicitaires par mois..
* Caractère Y : Ventes réalisées par mois.

**2**.- **Mise en ordre de la série double**

Elle consiste à organiser la suite des couples (xi,yi)i=1,…n de façon à éviter la redondance. Nous obtenons une nouvelle suite de couples ((xi,yj),nij)i=1,…r ; j=1,…,s avec

nij le nombre de couples de la série double qui représente la modalité xi du caractère X et la modalité yj du caractère Y

On a les relations suivantes :

1. **Tableau de Correspondance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Y  X | y1 | ………. | yj | ………….. | ys |
| x1 | n11 |  | n1j |  | n1s |
|  |  |  |  |  |  |
| xi | ni1 |  | nij |  | nis |
|  |  |  |  |  |  |
| xr | nr1 |  | nr1 |  | nrs |

n

1. **Tableau de Contingence**

Il est obtenu en remplaçant les effectifs nij par les fréquences fij= nij/n avec :

1. **Séries Marginales**

On peut définir, à partir de la série double deux séries marginales : celle de X et celle de Y

* La première ( de X) est obtenue en faisant la somme des effectifs (resp. des fréquences) de chaque ligne.

Elle est décrite par la suite finie ((xi, ni.)) ; i=1,…,r avec

* La deuxième (de Y) est obtenue en faisant la somme des effectifs (resp. des fréquences) de chaque colonne.

Elle est décrite par la suite finie ((yj, n.j)); j=1,…,s avec

1. **Séries Marginales Conditionnelles**

On peut définir, à partir de la série double deux séries marginales conditionnelles  :

* Celle de X conditionnée par Y et
* celle de Y conditionnée par X
* La première ( de X conditionnée par Y) est obtenue en fixant une colonne du tableau de correspondance ou de Contingence c’est à dire Y est fixée à la valeur yj.

Elle est définie par :

.

Elle est décrite par la suite finie ((xi, fi/j)) ; i=1,…,r avec

* La deuxième ( de Y conditionnée par X) est obtenue en fixant une ligne du tableau de correspondance ou de Contingence c’est à dire X est fixée à la valeur xi.

Elle est définie par :

.

Elle est décrite par la suite finie ((yj, fj/i)) ; j=1,…,s

avec

1. **Caractéristiques de la série double**

Elle est caractérisée par :

* Moyenne arithmétique de X =
* Moyenne arithmétique de Y =
* Variance de X :
* Variance de Y :
* Covariance de la série double

* Coefficient de corrélation linéaire

Pour répondre au premier objectif, on adopte la règle de décision suivante :

Règle De Décision

Si

Sinon

FinSi

Exemple traité

Supposons que la série double est définie par le tableau de correspondance suivant :

|  |  |  |  |
| --- | --- | --- | --- |
| X\Y | 10 | 20 | 50 |
| 1.17 | 20 | 10 | 40 |
| 1.25 | 15 | 25 | 10 |
| 3.15 | 5 | 10 | 15 |
| 4.67 | 4 | 10 | 12 |

1. Définir les séries marginales de X et de Y.
2. Définir les séries marginales conditionnelles

a.- de X quand Y=20.

b.- de Y quand X=3.15

1. Les deux caractères X et Y sont-ils corrélés ? Justifier !

EMSI

TD N° 3 -Séries Doubles

Exercice-1 \_\_\_\_

Pour mesurer l'impact d'une publicité, on observe un échantillon de 5 individus. A chaque individu, on présente l'affiche publicitaire en question et on lui demande de lui attribuer une note de 0 à 10. On mesure aussi le temps (en secondes) qu'il a passé à regarder cette affiche. On désigne

* **X** : La variable aléatoire Note attribuée à l'affiche ;
* **Y** : La variable aléatoire Temps passé à regarder l'affiche.

On obtient les 5 couples d'observations :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Valeur de **X** | 5 | 6 | 7 | 8 | 9 |
| Valeur de **Y** | 20 | 25 | 40 | 45 | 50 |

1. Les variables X et Y sont-elles corrélées ? Justifier !
2. La liaison linéaire entre X et Y est-elle Légitime ? Justifier !
3. a) Donner la droite de régression de Y par X :

y = A+ B.x

b) Prévoir la valeur de y pour x=15

1. a) Donner la courbe de régression de y par Ln(x) :

y= C+D.Ln(x)

b) Prévoir la valeur de y pour x=15.

1. Parmi les deux ajustements, quel est le plus Légitime ? Justifier

MOY(X)=(1.17X70+125x50+3.15x30+4.67x26)/176

MOY(Y)=(10x44+20x55+50x77)/176

S(X,Y)=(20x1.17x10+…+12x4.67x50)/176-moy(X).moy(Y)